Project 1: Kohonen maps on hand-written digits

Kristian Aurlien and Mateusz Paluchowski
EPFL, Switzerland

***Abstract*—This paper is part of the mini project on unsupervised and reinforcement learning in neural networks conducted at EPFL in fall of 2016. It's purpose is to present process of analyzing the effect of changing various parameters of Kohonen self organizing maps on the example MNIST dataset of hand-written digits.**

I. Introduction

Self organizing maps (SOM) is a kind of Artificial Neural Network, where competitive unsupervised learning is used to transform a high dimensional (continuous) input space into a low dimensional (typically 1 or 2), discrete output space. Kohonen maps is a topology-preserving specification of SOM, with strong biological and mathematical background. Kohonen networks learn to create maps of the input space in a self-organizing way. In this paper we look at how different parameters (learning rate, map size, neighborhood function) affects the convergence and performance of Kohonen maps.

II. Main Part

A. Baseline

Just as it was described in the report guidelines, we start with a baseline Kohonen network of 6x6 neurons which are arranged on a square grid with unit distance and use a Gaussian neighborhood function with constant standard deviation $\sigma = 3$. As mentioned by S. Haykin in Neural Networks and Learning Machines[1] the general proof of SOM convergence has been proven only for one-dimensional case and when it comes to convergence of higher-dimensional cases, general solution remains to be found. At first we tried our own convergence criterion based on following rule:

$$\frac{\|centers_{n+1} - centers_n\|}{\|centers_n\|} < 0.04$$

where 0.04 was picked heuristically just for the baseline network. However, we didn't use it due to fact that the algorithm might be terminated too soon and we might end up in meta-stable state.

Following Haykin's guidelines we assumed that the ordering phase takes on avarege 1000 iterations, whilst the convergence phase must be at least 500 times the number of neurons in the network. Thus, for this Kohonen network we ended up with doing 19,000 iterations in order to achieve stable state.

As visualized in the Figure 1, although SOM is in the stable state, the results are far from perfect. Prototypes of the digits are almost indistinguishable. However such outcome shouldn't be surprising considering what values were passed as parameters to the Kohonen network SOM.

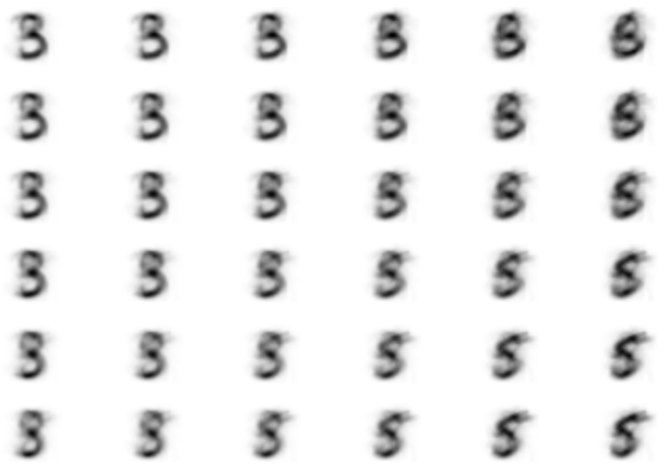

Fig. 1. The visualization of achieved SOM prototypes after 19k iterations for 6x6 Kohonen network with neighborhood function consisting of constant $\sigma = 3$.

Figure 1 shows the result of using a large parameter $\sigma = 3$ which, for 6x6 network, covers almost all prototypes and affects them in each iteration. Moreover, running 19000 iterations with a constant σ results in averaging of the prototypes. There is only minimal difference between prototypes and all of them are a combination of original [3 5 6 8] set which was used for training.

B. Automatic label assignment

Our label assignments match each prototype in the SOM with the label of the closest data-points from our training data. We do the assignment using a K-nearest-neighbors classifier, using K=3. The choice of K is based on empirical testing. Though this is a simple model, it seems to give us reasonably good matches. Assignments achieved for the prototypes generated by Kohonen SOM with exponential decay of neighborhood function (Figure 2) is presented in the Figure 3.

C. Kohonen map size exploration

To see how the size of the map affects the learning, we explore two sets of Kohonen maps. First, with constant $\sigma = 3$ and three different dimensions: 8x8, 10x10 and 12x12. Second, with constant 8x8 dimension and three different sigmas: 1, 3, 5. Achieved results backs up our previous remarks about influence of the size of neighborhood - it has great influence on the final shape of the prototypes visualized in Appendix A as follows:

- Figure 5 - 8x8, $\sigma = 1$

- Figure 6 - 8x8, $\sigma = 3$
- Figure 7 - 8x8, $\sigma = 5$
- Figure 8 - 10x10, $\sigma = 3$
- Figure 9 - 12x12, $\sigma = 3$

Our size exploration shows a clear trend when it comes to the role of the width of neighborhood function. The bigger the size, the more SOM centers are affected at each step. This results in bigger averaging over larger amount of unequal neighboring centers and thus prototypes become less accurate. Both our exploration sets seem to confirm this relation - prototypes become more informative when either we increase map dimension for constant $\sigma = 3$ or decrease σ for constant dimension 8x8. From these results we see that the optimal width of the neighborhood function does depend on the size of Kohonen map.

Figure 5 presents our best results from the two explored sets. For more details we encourage the reader to see included jupyter notebook which should be treated as additional appendix for this report.

D. Neighborhood exponential decay

Using a gradually decreasing neighborhood function, has shown to be a good strategy. Inspired by Haykin, we initialize the neighborhood with a "width" so that it covers most of the map, and gradually decrease it down to zero.

$$\sigma(n) = \sigma_0 \exp(-\frac{n}{\tau_1}), n = 0, 1, 2, ...$$

The wanted behavior is than obtained by using the following parameters, d being the dimension of the map:

$$\sigma_0 = \sqrt{2 * (\frac{d}{2})^2}$$

$$\tau_1 = \frac{1000}{\log \sigma_0}$$

With the neighborhood exponential decay, we obtain the following map:

As we can see in the Figure 2 the results are much better then what we have achieved before. Most, but not all, of the prototypes are clear and moreover all of the numbers from original [3 5 6 8] set can be found in the prototypes.

E. Learning rate exponential decay

Similar to the decreasing the neighborhood size over time, we also try using learning rate with exponential decay. We want the learning rate to be relatively high in the beginning (0.1), and then decrease to 0.01 as we get closer to convergence. The decaying learning rate is given by

$$\eta(n) = \eta_0 \exp{(\frac{-n}{\tau_2})}$$

Using $\eta_0 = 0.1$, $n_{max} = 19000$, $\eta(n_{max}) = 0.01$ and solving for τ_2, we get that $\tau_2 = 82152$ satisfies these requirements. With the decreasing neighborhood function and decreasing learning rate, we finally end up with the following model:

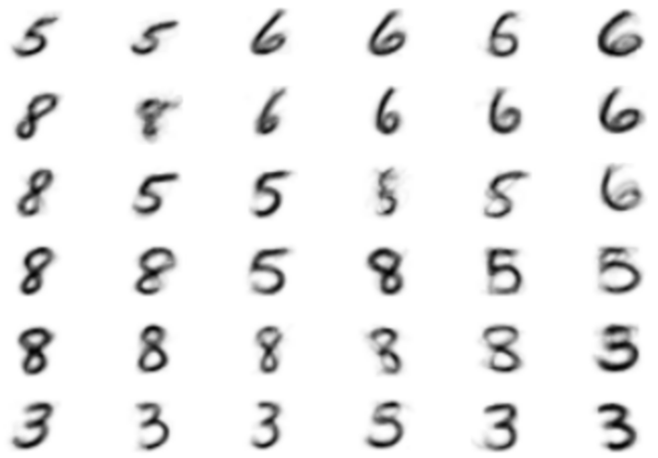

Fig. 2. The visualization of achieved SOM prototypes after **19k** iterations for **6x6** Kohonen network with exponential decay of neighborhood function for which $\boldsymbol{\sigma_0 = 4.24}$.

5	5	6	6	6	6
8	8	6	6	6	6
8	5	5	5	5	6
8	8	5	8	5	3
8	8	8	3	8	5
3	3	3	5	3	3

Fig. 3. Generated assignments for Fig2

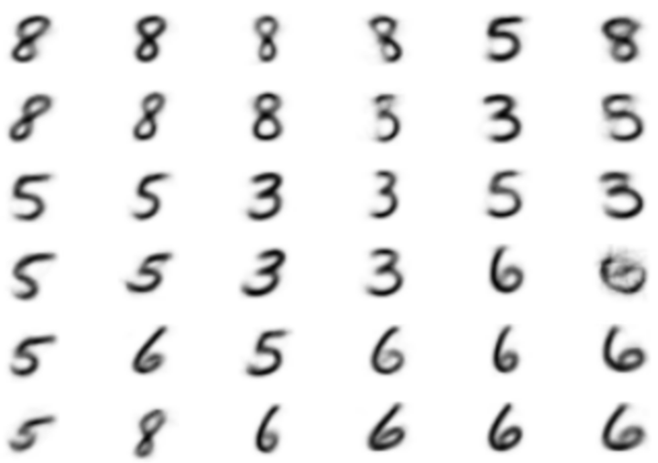

Fig. 4. The visualization of achieved SOM prototypes after **19k** iterations for **6x6** Kohonen network with exponential decay of neighborhood function for which $\boldsymbol{\sigma_0 = 4.24}$ and exponential decay of learning rate for which $\boldsymbol{\eta_0 = 0.1}$.

This is by far the best result we achieved. All of the prototypes are clear and its labels are easily distinguishable, while they also grasp subtle differences in the way one digit might be hand-written.

III. Conclusion

Achieving stable state in case of multidimensional Kohonen Self Organizing Maps is not an easy task. Lots of different factors play major role when it comes to correct convergence process and avoiding meta-stable state. We found that, especially in the case of small network dimensions, the size of the selected neighborhood greatly matters. Proper initialization of learning rate is also crucial. Last but not least, decreasing both of the aforementioned factors during the process of learning is beneficial for the final outcome. However, this decrease cannot be too strong, otherwise the network could be 'frozen' too soon and meta-stable state without actual convergence to optimal solution would be obtained.

REFERENCES

[1] Simon S. Haykin. 2009. Neural networks and learning machines, Prentice Hall

[2] Teuvo Kohonen. 1997. Self-Organizing Maps. (Ed.). Springer-Verlag New York, Inc., Secaucus, NJ, USA.

[3] Ral Rojas. 1996. Neural Networks: A Systematic Introduction. Springer-Verlag New York, Inc., New York, NY, USA

APPENDIX

Vizualization of SOM prototypes for Kohonen map size expploration.

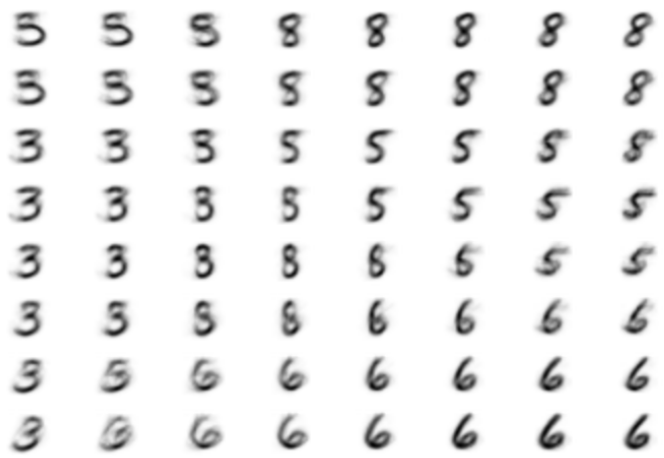

Fig. 5. The visualization of achieved SOM prototypes after **33k** iterations for **8x8** Kohonen network with neighborhood function consisting of constant $\sigma = \mathbf{1}$.

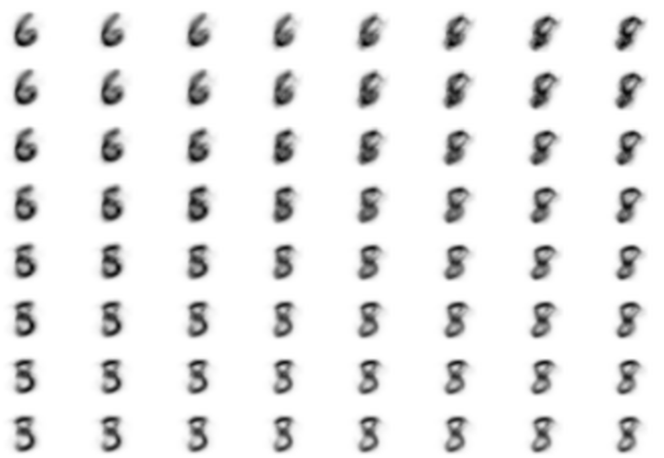

Fig. 6. The visualization of achieved SOM prototypes after **33k** iterations for **8x8** Kohonen network with neighborhood function consisting of constant $\sigma = \mathbf{3}$.

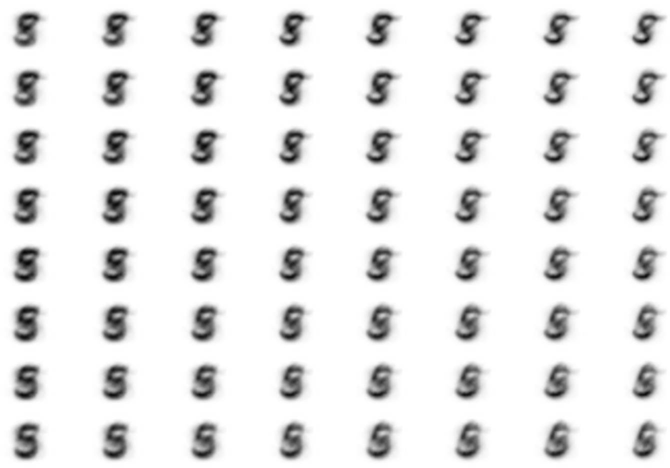

Fig. 7. The visualization of achieved SOM prototypes after **33k** iterations for **8x8** Kohonen network with neighborhood function consisting of constant $\sigma = \mathbf{5}$.

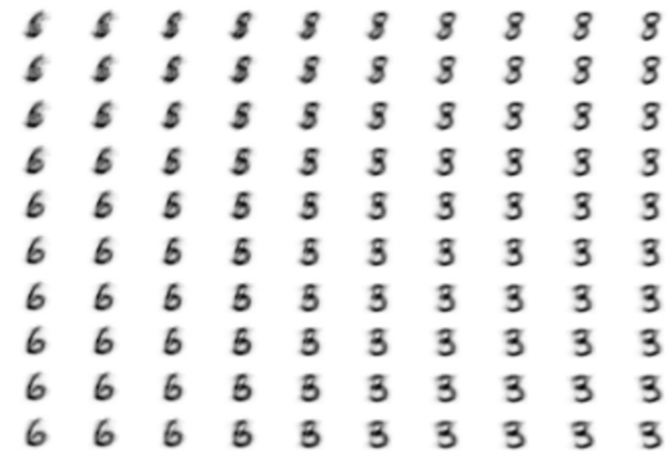

Fig. 8. The visualization of achieved SOM prototypes after **51k** iterations for **10x10** Kohonen network with neighborhood function consisting of constant $\sigma = \mathbf{3}$.

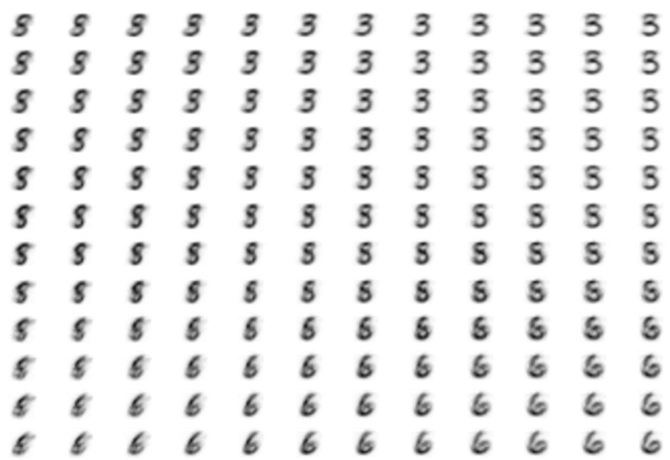

Fig. 9. The visualization of achieved SOM prototypes after **73k** iterations for **12x12** Kohonen network with neighborhood function consisting of constant $\sigma = \mathbf{3}$.